

Téma 2.: Bodové a intervalové rozložení četností

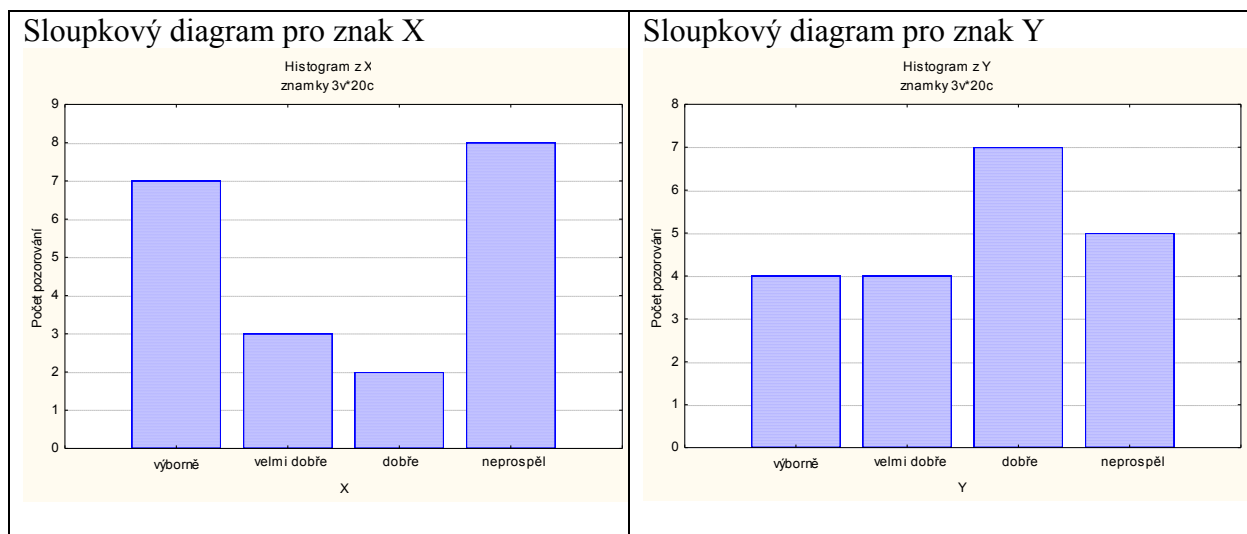
Nejprve budeme pracovat s datovým souborem znamky.sta, který obsahuje údaje o známkách z matematiky, angličtiny a pohlaví 20 studentů 1. ročníku (viz skripta Popisná statistika, příklad 2.4).

Úkol 1.: Vytvořte

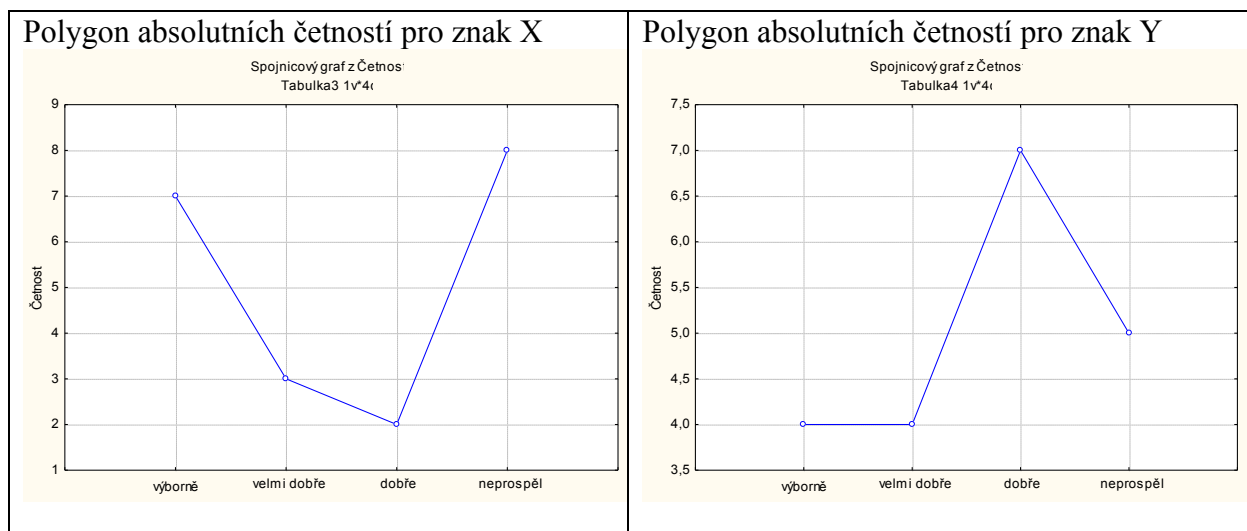
- sloupkový diagram absolutních četností znaků X a Y,
- polygon absolutních četností znaků X a Y
- graf četnostní funkce znaku X,
- graf empirické distribuční funkce znaku X.

Návod:

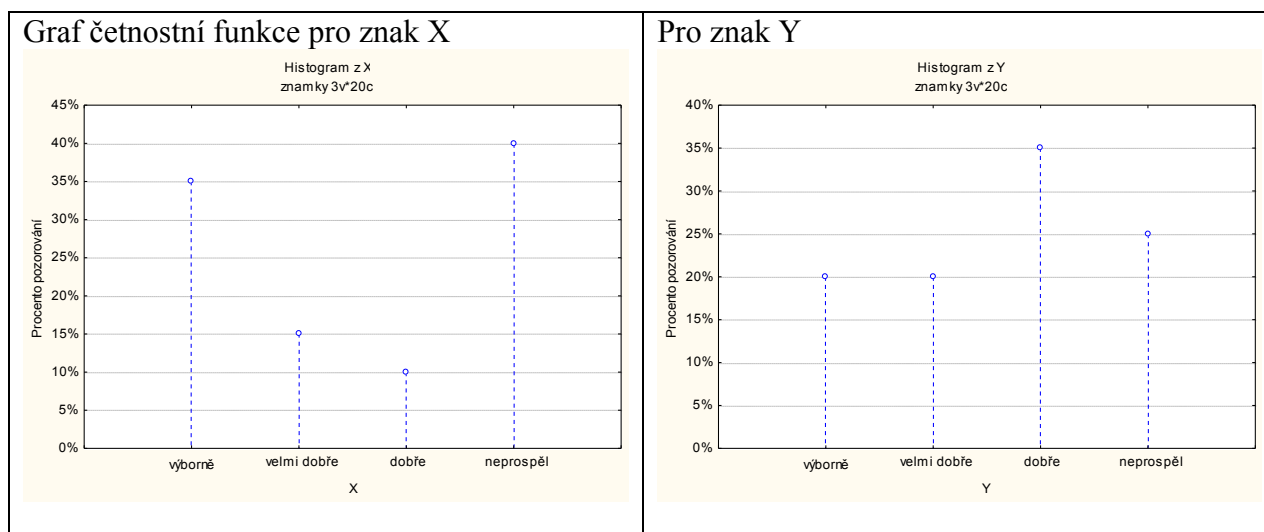
ad a) Grafy – Histogramy – Proměnné X, Y – OK- vypneme Normální proložení – Details – zaškrtneme Mezery mezi sloupci - OK.



ad b) V pracovním sešitu vstoupíme do tabulky rozložení četností proměnné X resp. Y. Nastavíme se na řádek označený ChD. Pomocí Případy – Odstranit vymažeme tento řádek. Nastavíme se kurzorem na Četnost - klikneme pravým tlačítkem – Grafy bloku dat – Spojnicový graf: celé sloupce. Vykreslí se polygon absolutních četností.

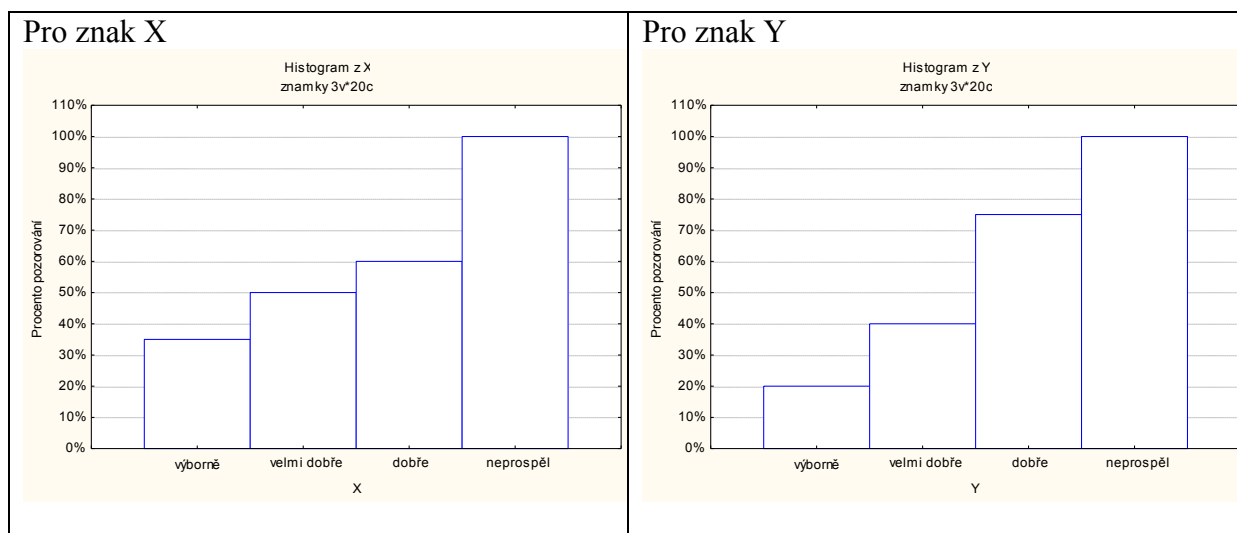


ad c) Při tvorbě histogramu vypneme Normální proložení, zadáme v Detailech Osa Y % - 2 x klikneme myší na pozadí grafu – vybereme Graf: Obecné – zaškrtneme Značky – vybereme Graf: Sloupce – Typ: Čáry, nastavíme čárkovanou čáru.



ad d) Při tvorbě histogramu vypneme Normální proložení, zadáme v Detailech volbu Zobrazovaný typ: Kumulativní, Osa Y % - 2x klikneme myší na pozadí grafu – vybereme Graf: Sloupce – Typ: Obdélníky.

Upozornění: V tomto grafu se objeví svislé čáry, které samozřejmě do grafu empirické distribuční funkce nepatří.



Úkol 2.: Vytvořte variační řady známek z matematiky a angličtiny pouze

- pro ženy,
- pro muže.

Návod:

ad a) Statistiky – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné X, Y – OK – vybereme Select Cases - zaškrtneme Zapnout filtr – do okénka některé, vybrané pomocí výrazu zapíšeme $Z = 0$, OK, Výpočet.

Variační řada známek z matematiky pro ženy:

Tabulka četností: X: známka z matematiky (znamky) Zhrnout podmínku: Z=0				
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
výborně	5	5	50,00000	50,0000
velmi dobře	2	7	20,00000	70,0000
dobře	1	8	10,00000	80,0000
neprospěl	2	10	20,00000	100,0000
ChD	0	10	0,00000	100,0000

Variační řada známek z angličtiny pro ženy:

Tabulka četností: Y: známka z angličtiny (znamky) Zhrnout podmínku: Z=0				
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
výborně	4	4	40,00000	40,0000
velmi dobře	2	6	20,00000	60,0000
dobře	1	7	10,00000	70,0000
neprospěl	3	10	30,00000	100,0000
ChD	0	10	0,00000	100,0000

ad b) Statistiky – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné X, Y – OK – vybereme Select Cases - zaškrtneme Zapnout filtr – do okénka některé, vybrané pomocí výrazu zapíšeme $Z = 1$, OK, Výpočet.

Variační řada známek z matematiky pro muže:

Tabulka četností: X: známka z matematiky (znamky) Zhrnout podmínku: Z=1				
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
výborně	2	2	20,00000	20,0000
velmi dobře	1	3	10,00000	30,0000
dobře	1	4	10,00000	40,0000
neprospěl	6	10	60,00000	100,0000
ChD	0	10	0,00000	100,0000

Variační řada známek z angličtiny pro muže:

Tabulka četností: Y: známka z angličtiny (znamky) Zhrnout podmínku: Z=1				
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
velmi dobře	2	2	20,00000	20,0000
dobře	6	8	60,00000	80,0000
neprospěl	2	10	20,00000	100,0000
ChD	0	10	0,00000	100,0000

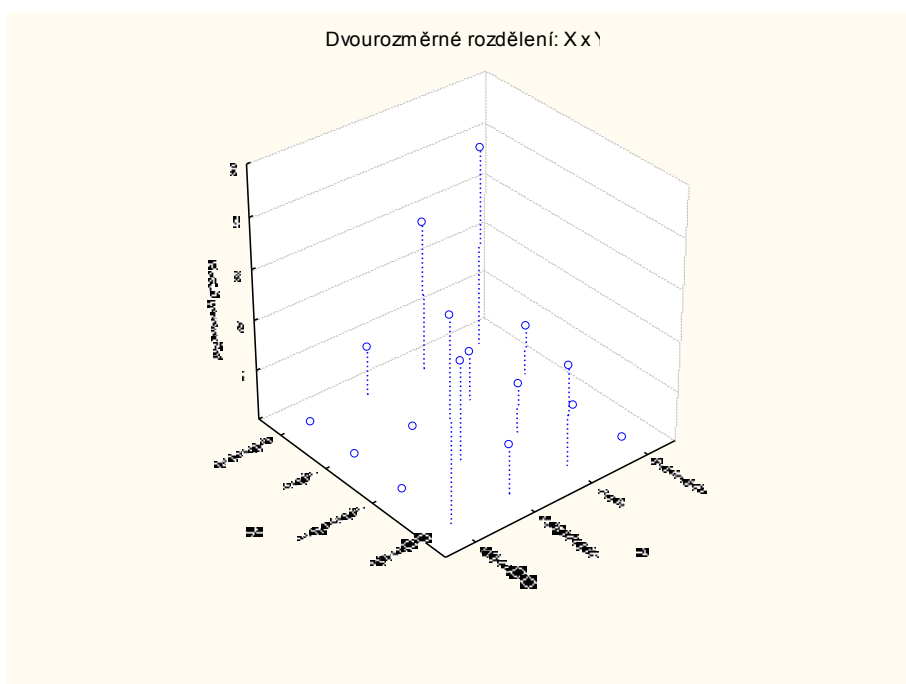
Úkol 3.: Nadále budeme pracovat s celým datovým souborem. Vytvoříme kontingenční tabulku simultánních absolutních četností znaků X a Y a graf simultánní četnostní funkce.

Návod: Statistiky – Základní statistiky/tabulky – odškrtneme Zapnout filtr – OK - Kontingenční tabulky – OK – Select cases — Specif. tabulky - List 1 X, List 2 Y, OK, Výpočet.

Kontingenční tabulka (znamky)					
Četnost označených buněk > 10					
(Marginální součty nejsou označeny)					
X	Y	Y	Y	Y	Řádk. součty
	výborně	velmi dobře	dobře	neprospěl	
výborně	4	1	2	0	7
velmi dobře	0	2	1	0	3
dobře	0	0	1	1	2
neprospěl	0	1	3	4	8
Vš.skup.	4	4	7	5	20

Vidíme, že ve výběrovém souboru byly 4 studenti, kteří měli z obou předmětů „výborně“, jeden student, který měl z matematiky „výborně“ a z angličtiny „velmi dobře“ atd. až 4 studenti, kteří z obou předmětů neprospěli.

Vytvoření grafu simultánní četnostní funkce: Na liště aktivujeme Výsledky: kontingenční tabulky – Details - 3D histogramy. Vzniklý graf je třeba upravit: 2x klikneme myší na pozadí grafu – Rozvržení grafu – Typ – Špičky – OK.



Upozornění: Ve Statistice verze 6 je třeba ručně měnit nastavení os X a Y. 2x klikneme myší na pozadí grafu – Osa: Měřítka – Mód: Ručně – Minimum 0 (a totéž provedeme pro Osu Y).

Graf lze natáčet pomocí Zorného bodu.

Úkol 4.: Vytvořte kontingenční tabulku sloupcově a řádkově podmíněných relativních četností znaků X a Y.

Návod: Aktivujeme na liště Výsledky: kontingenční tabulky – Možnosti - zaškrtneme ve sloupci Výpočet tabulek volbu Procenta z počtu ve sloupci (resp. Procenta z počtu v řádku) – Výpočet.

Kontingenční tabulka sloupcově podmíněných relativních četností :

Kontingenční tabulka (znamky) Četnost označených buněk > 10 (Marginální součty nejsou označeny)						
	X	Y	Y	Y	Y	Řádk. součty
		výborn ě	velmi dobře	dobře	neprospěl	
Četnost	výborr ě	4	1	2	0	7
Sloupc. četn.		100,00%	25,00%	28,57%	0,00%	
Četnost	velmi dobře	0	2	1	0	3
Sloupc. četn.		0,00%	50,00%	14,29%	0,00%	
Četnost	dobře	0	0	1	1	2
Sloupc. četn.		0,00%	0,00%	14,29%	20,00%	
Četnost	neprospěl	0	1	3	4	8
Sloupc. četn.		0,00%	25,00%	42,86%	80,00%	
Četnost	Vš.skup.	4	4	7	5	20

Interpretace např. 4. řádku ve 2. sloupci: V souboru byli 4 studenti, kteří měli velmi dobře z angličtiny. Mezi nimi byl jeden, který neprospěl z matematiky, což představuje $1/4 = 25\%$.

Kontingenční tabulka řádkově podmíněných relativních četností:

Kontingenční tabulka (znamky) Četnost označených buněk > 10 (Marginální součty nejsou označeny)						
	X	Y	Y	Y	Y	Řádk. součty
		výborn ě	velmi dobře	dobře	neprospěl	
Četnost	výborr ě	4	1	2	0	7
Řádk. četn.		57,14%	14,29%	28,57%	0,00%	
Četnost	velmi dobře	0	2	1	0	3
Řádk. četn.		0,00%	66,67%	33,33%	0,00%	
Četnost	dobře	0	0	1	1	2
Řádk. četn.		0,00%	0,00%	50,00%	50,00%	
Četnost	neprospěl	0	1	3	4	8
Řádk. četn.		0,00%	12,50%	37,50%	50,00%	
Četnost	Vš.skup.	4	4	7	5	20

Interpretace např. 2. sloupce ve 4. řádku: V souboru bylo 8 studentů, kteří neprospěli z matematiky. Mezi nimi byl jeden, který měl velmi dobře z angličtiny, což představuje $1/8 = 12,5\%$.

Nyní se budeme věnovat datovému souboru ocel.sta. Obsahuje údaje o mezi plasticity a mezi pevnosti 60 vzorků oceli (viz skripta Popisná statistika, příklad 2.13).

Úkol 4.: Načteme soubor ocel.sta. Proměnným X a Y vytvoříme návěští „mez plasticity“ a „mez pevnosti“. Podle Sturgersova pravidla najdeme optimální počet třídících intervalů pro znaky X a Y a vhodně stanovíme meze třídících intervalů.

Návod: Soubor – Otevřít – vybereme příslušný adresář se souborem ocel.sta – Otevřít. Kurzor nastavíme na X – 2x klikneme myší – Dlouhé jméno mez plasticity – OK, kurzor nastavíme na Y – 2x klikneme myší – Dlouhé jméno mez pevnosti – OK.

Protože případů je 60, podle Sturgersova pravidla je optimální počet třídících intervalů 7. Musíme zjistit minimum a maximum, abychom vhodně stanovili třídící intervaly: Statistika - Základní statistika/tabulky – Popisné statistiky - OK - Proměnné X,Y – OK – Detailní výsledky – ponecháme zaškrtnuté Minimum&maximum – Výpočet.

Pro X je minimum 33 a maximum 160, tedy dolní mez prvního třídícího intervalu volíme 30, horní mez posledního třídícího intervalu 170. Celkem tedy třídící intervaly pro znak X budou: (30,50>, (50,70>, (70,90>, (90,110>, (110,130>, (130,150>, (150,170>

Pro Y je minimum 52 a maximum 189, tedy dolní mez prvního třídícího intervalu volíme 50, horní mez posledního třídícího intervalu 190. Celkem tedy třídící intervaly pro znak Y budou: (50,70>, (70,90>, (90,110>, (110,130>, (130,150>, (150,170>, (170,190>.

Úkol 5.: Provedeme zakódování hodnot proměnných X a Y do příslušných třídících intervalů. Všem hodnotám proměnné X, které leží v intervalu (30,50>, přiřadíme hodnotu 1 atd. až všem hodnotám proměnné X, které leží v intervalu (170,190>, přiřadíme hodnotu 7. Analogicky pro Y.

Návod: Vytvoříme dvě nové proměnné: Vložit – Přidat proměnné – 2 – Za Y – OK – přejmenujeme je na RX a RY. Nastavíme se kurzorem na RX – Data – Překódovat - vyplníme podmínky pro všech 7 kategorií. (Pozor – podmínky píšeme ve tvaru $X > 30$ and $X \leq 50$ atd.). Pak klepneme na OK.

Překódovat hodnoty proměnné 3: RX

Kategorie 1
Zahmout pokud: Zahmout pokud:
 $X > 30$ and $X \leq 50$
Nová hodnota 1
 hodnota: 1
 Kód ChD

Kategorie 2
Zahmout pokud: Zahmout pokud:
 $X > 50$ and $X \leq 70$
Nová hodnota 2
 hodnota: 2
 Kód ChD

Kategorie 3
Zahmout pokud: Zahmout pokud:
 $X > 70$ and $X \leq 90$
Nová hodnota 3
 hodnota: 3
 Kód ChD

Kategorie 4
Zahmout pokud: Zahmout pokud:
 $X > 90$ and $X \leq 110$
Nová hodnota 4
 hodnota: 4
 Kód ChD

OK
Storno
Zrušit vše
Otevřít...
Uložit jako...
Proměnná...

Jiné
Pokud nejsou splněny žádné podmínky, nastavit hodnoty na:
 Kód ChD
 hodnota:
 nezměněna

Analogicky překódujeme hodnoty proměnné Y do proměnné RY.

Překódovat hodnoty proměnné 4: RY

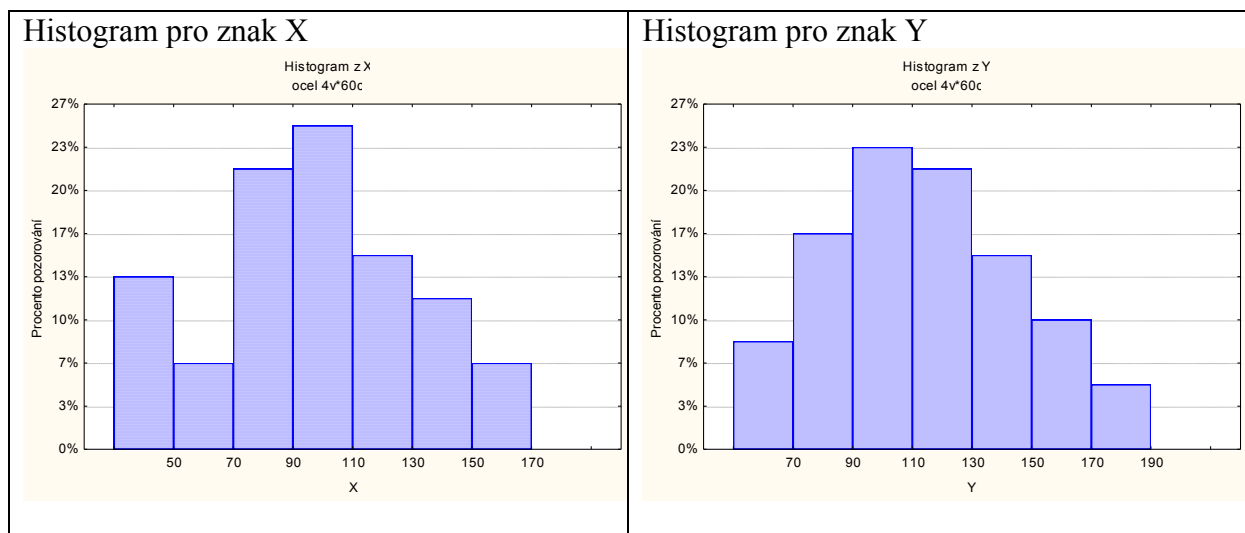
Kategorie 1 Zahrnout pokud: $Y > 50 \text{ and } Y \leq 70$	Nová hodnota 1 <input checked="" type="radio"/> hodnota: 1 <input type="radio"/> Kód ChD
Kategorie 2 Zahrnout pokud: $Y > 70 \text{ and } Y \leq 90$	Nová hodnota 2 <input checked="" type="radio"/> hodnota: 2 <input type="radio"/> Kód ChD
Kategorie 3 Zahrnout pokud: $Y > 90 \text{ and } Y \leq 110$	Nová hodnota 3 <input checked="" type="radio"/> hodnota: 3 <input type="radio"/> Kód ChD
Kategorie 4 Zahrnout pokud: $Y > 110 \text{ and } Y \leq 130$	Nová hodnota 4 <input checked="" type="radio"/> hodnota: 4 <input type="radio"/> Kód ChD

Jiné
Pokud nejsou splněny žádné podmínky, nastavit hodnoty na:
 Kód ChD
 hodnota:
 nezměněna

OK
Storno
Zrušit vše
Otevřít...
Uložit jako...
Proměnná...

Úkol 6.: Vytvoříme histogram pro X a pro Y.

Návod: Grafy – Histogramy – Proměnné X – vypneme Normální proložení – Details – zaškrtneme Hranice – Určit hranice – 50 70 90 110 130 150 170 OK – Osa Y %. Po vykreslení histogramu lze 2 x klepnout na pozadí grafu a ve volbě Všechny možnosti měnit různé vlastnosti grafu.

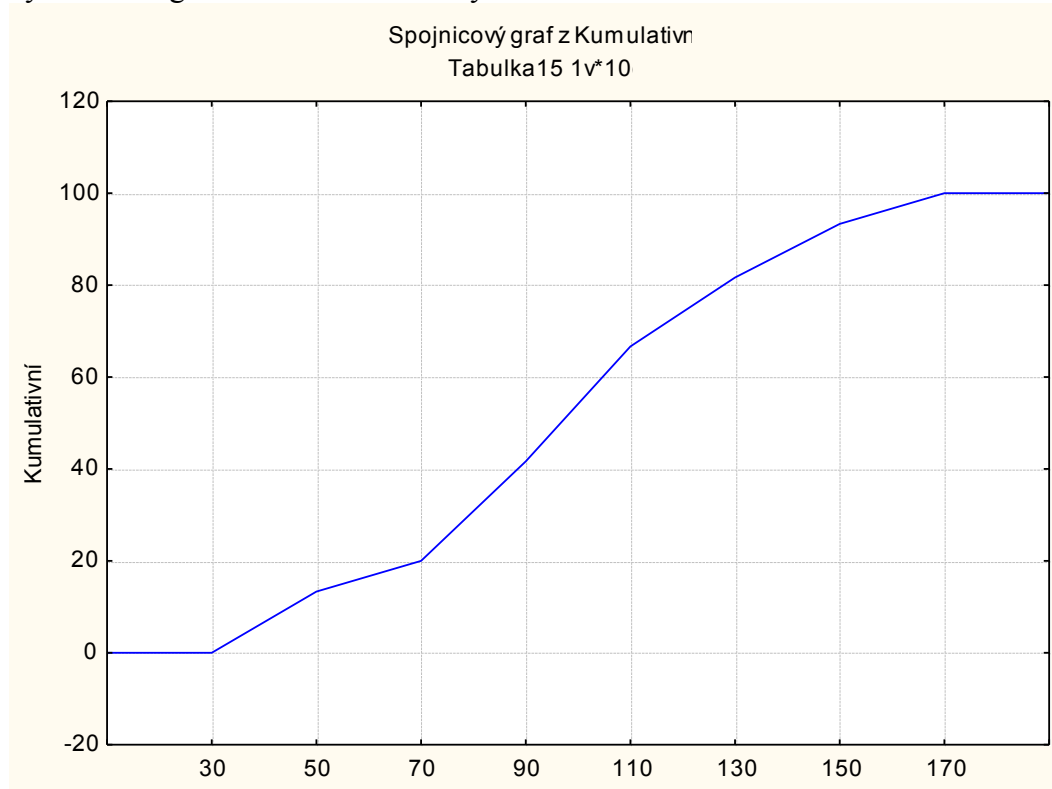


Úkol 7.: Vytvoříme graf intervalové empirické empirické distribuční funkce pro X.

Návod: Vytvoříme tabulku četností pro proměnnou RX. Před 1. případ vložíme dva řádky, u nichž do Kumulativní rel. četnost napíšeme 0. Do sloupce Kategorie napíšeme 10, 30, 50, ..., 190:

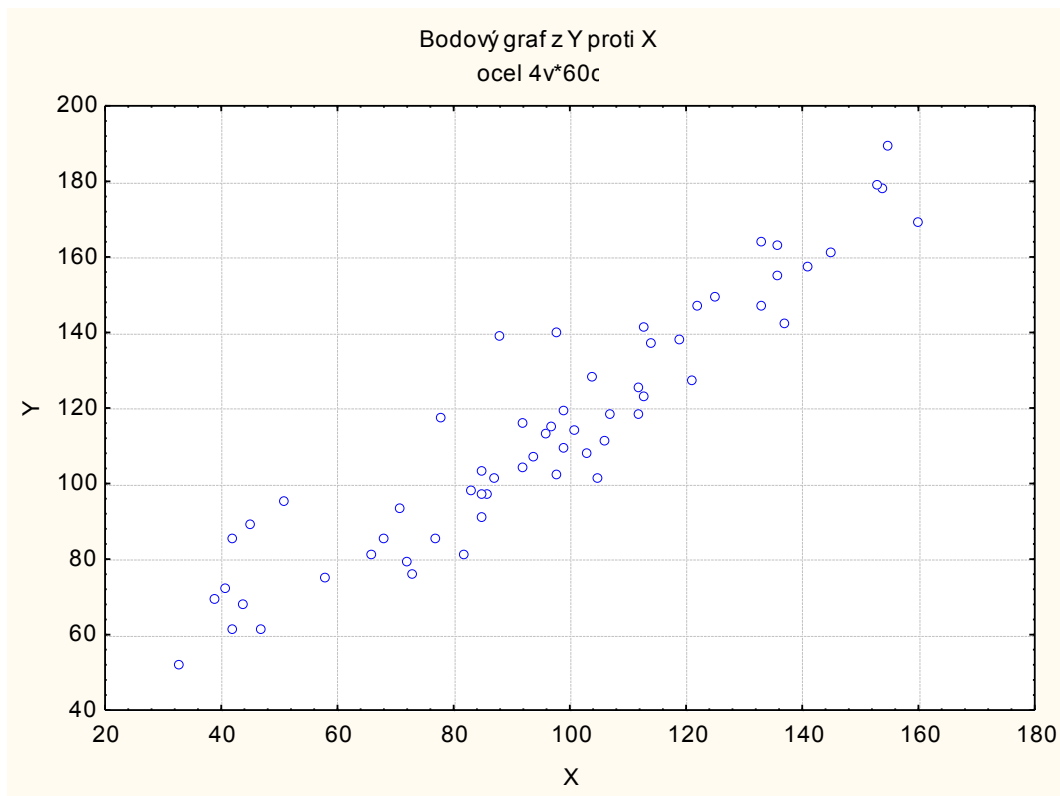
Kategorie	Tabulka četností:RX (ocel)			
	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
10				0,0000
30				0,0000
50	8	8	13,33333	13,3333
70	4	12	6,66667	20,0000
90	13	25	21,66667	41,6667
110	15	40	25,00000	66,6667
130	9	49	15,00000	81,6667
150	7	56	11,66667	93,3333
170	4	60	6,66667	100,0000
190	0	60	0,00000	100,0000

Nastavíme se kurzorem na Kumulativní rel. četnost – klikneme pravým tlačítkem – Grafy bloku dat – Vlastní graf bloku podle sloupce – Sloupcové grafy (Proměnné) – OK. Ve vytvořeném grafu odstraníme značky a změníme rozsah hodnot na vodorovné ose od 1 do 10.



Úkol 8.: Nakreslíme dvourozměrný tečkový diagram pro (X,Y).

Návod: Grafy – Bodové grafy – Proměnné X,Y – OK - vypneme Lineární proložení – OK.



Vidíme, že mezi oběma proměnnými existuje určitý stupeň přímé lineární závislosti – s růstem hodnot meze plasticity vesměs rostou hodnoty meze pevnosti a naopak.

Práce se systémem SPSS

Bodové zpracování četností

Otevřeme datový soubor znamky.sav. Vytvoření návěští jednotlivým proměnným a popis jejich variant se provádí v okně Variable View.

Tabulku rozložení četností pro známky z matematiky získáme takto: Analyze – Descriptive Statistics – Frequencies – Variable X – OK. Dostaneme tabulku rozložení četností.

Pro grafické znázornění četností do ni vstoupíme dvojklikem. Vezmeme do bloku absolutní četnosti variant 1, 2, 3, 4. Klikneme pravým tlačítkem – Create Graph – Bar. Spojnicový graf či výsečový diagram získáme tak, že v Create Graph vybereme Line resp. Pie.

Vlastnosti grafu se dají měnit v Chart Editor, do něhož vstoupíme tak, že 2x klikneme myší na vytvořený graf.

Graf četnostní funkce získáme tak, že v tabulce rozložení četností se nastavíme na Percent a v Create Graph vybereme Dot.

Graf empirické distribuční funkce získáme tak, že tabulce rozložení četností se nastavíme na Cumulative Percent a v Create Graph vybereme Bar. Pak vstoupíme do Chart Editoru a změníme šířku sloupce na 100%.

Chceme-li pracovat pouze s částí datového souboru (např. pouze se známkami z matematiky pro muže), jdeme do menu Data – Select Cases – If condition is satisfied – If – Z=1 – Continue – OK.

Vytvoření kontingenční tabulky

Kontingenční tabulku simultánních absolutních četností známek z matematiky a angličtiny vytvoříme tímto způsobem: Analyze – Descriptive Statistics – Crosstabs – Row(s) X, Column(s) Y – OK.

Chceme-li získat simultánní relativní četnosti nebo řádkově či sloupcově podmíněné relativní četnosti, v Crosstabs zvolíme Cells a zaškrtneme Percentages Total resp. Row resp. Column. Složený sloupcový diagram dostaneme tak, že v Crosstabs zaškrtneme Display clustered charts.

Intervalové zpracování četností

Otevřeme datový soubor ocel.sav.

V úkolu 4 jsme zjistili, že hodnoty znaku X (tj. meze plasticity) je vhodné rozřídít do 7 třídicích intervalů: (30,50>, (50,70>, (70,90>, (90,110>, (110,130>, (130,150>, (150,170> Provedeme zakódování hodnot znaku X do těchto třídicích intervalů.

Transform – Recode into Different Variables – Numeric Variable X, Output variable RX – Change – Old and new values – v Old Value zaškrtneme Range, 30 through 50, v New Value zadáme Value 1 – Add, vrátíme se do Old Value, Range 51 through 70, v New Value zadáme value 2 – Add atd.

Podobně zakódujeme hodnoty znaku Y do proměnné RY.

Vytvoření histogramu pro X:

Graphs – Legacy Dialogs – Histogram – Variable X – OK.

Vzniklý histogram upravíme v Chart Editor: Options – Un-Bin Element – Binning – na X Axis zaškrtneme Custom, Interval width 20, zaškrtneme Custom value from anchor 30 – Apply.

Nakreslení grafu intervalové empirické distribuční funkce pro X:

Vytvoříme tabulku rozložení četností pro znak X. V této tabulce se nastavíme na Cumulative Percent a v Create Graph vybereme Line.

Nakreslení dvourozměrného tečkového diagramu pro (X, Y):

Graphs – Legacy Dialogs – Scatter/Dot – Define – Y Axis Y, X Axis X – OK.